# Estimation of slope for linear regression model with uncertain prior information and Student-t error

**Shahjahan Khan**
Department of Mathematics and Computing
Australian Centre for Sustainable Catchments
University of Southern Queensland
Toowoomba, Queensland, Australia
*Email: khans@usq.edu.au*
and
**A.K.Md.E. Saleh**
School of Mathematics and Statistics
Carleton University, Ottawa, Canada

**Abstract**

This paper considers estimation of the slope parameter of the linear regression model with Student-t errors in the presence of uncertain prior information on the value of the unknown slope. Incorporating uncertain non-sample prior information with the sample data the unrestricted, restricted, preliminary test, and shrinkage estimators are defined. The performances of the estimators are compared based on the criteria of unbiasedness and mean squared errors. Both analytical and graphical methods are explored. Although none of the estimators is uniformly superior to the others, if the non-sample information is close to its true value, the shrinkage estimator over performs the rest of the estimators.

**Keywords and Phrases**: Multiple regression model; Student-t errors; preliminary test and shrinkage estimators; bias, mean square error and relative efficiency; mixture distribution of normal and inverted gamma; non-central chi-square and F distributions; and incomplete beta ratio.

**AMS 2000 Subject Classification**: Primary 62F30, Secondary 62H12 and 62F10.

## 1 Introduction

Customarily the classical estimators of unknown parameters are based exclusively on the sample data. Such estimators disregard any other kind of non-sample prior information in its definition. The notion of inclusion of non-sample information to the estimation of parameters has been introduced to 'improve' the quality of the estimators. The natural expectation is that the inclusion of additional information would result in a better estimator. In some cases this may be true, but in many other cases the risk of worse consequences can not be ruled out. A number of estimators have been introduced in the literature that, under particular situation, over performs the traditional exclusive sample data based unbiased estimators when judged by criteria such as the mean squared error and squared error loss function.

In the wake of increasing criticism on the inappropriate use of the normal distribution to model the errors there is a growing trend to use, often more appropriate, Student-t model. Fisher (1956, p.133) warned against the consequences of inappropriate use of the traditional normal model. Fisher (1960, p.46) analyzed Darwin's data (cf. Box and Taio, 1992, p. 133) by using a non-normal model. Fraser and Fick (1975) analyzed the same data by the Student-t model. Zellner (1976) provided both Bayesian and frequentist analyses of the multiple regression model with Student-t errors. Fraser (1979) illustrated the robustness of the Student-t model. Prucha and Kelegian (1984) proposed an estimating equation for the simultaneous equation model with the Student-t errors. Ullah and Walsh (1984) investigated the optimality of different types of tests used in econometric studies for the multivariate Student-t model. The interested readers may refer to the more recent work of Singh (1988), Lange et al. (1989), Giles (1991), Khan (1992), Anderson (1993), Spanos (1994), and Khan (1998) for different applications of the Student-t models. For a wide range of applications of the Student-t models refer to Lange et al. (1989).

There has been many studies in the area of the 'improved' estimation following the seminal work of Bancroft (1944) and later Han and Bancroft (1968). They developed the preliminary test estimator that uses uncertain non-sample prior information (not in the form of prior distributions), in addition to the sample information. Stein (1956) introduced the Stein-rule (shrinkage) estimator for multivariate normal population that dominates the usual maximum likelihood estimators under the squared error loss function. In a series of papers Saleh and Sen (1978, 1985) explored the preliminary test approach to Stein-rule estimation. Many authors have contributed to this area, notably Sclove et al. (1972), Judge and Bock (1978), Stein (1981), Maatta and Casella (1990), and Khan (1998), to mention a few. Khan and Saleh (1995, 1997) investigated the problem for a family of Student-t populations. However, the relative performance of the preliminary test and shrinkage estimators of the slope parameter of linear regression model with Student-t error has not been investigated.

It is well known that the mle of the slope parameter is unbiased. We wish to search for an alternative estimator of the slope parameter that is biased but may well have some superior statistical property in terms of another more popular statistical criterion, namely the mean square error. In this process, we define three biased estimators: the restricted estimator (RE) with a *coefficient of distrust*, the preliminary test estimator (PTE) as a linear combination of the mle and the RE, and the shrinkage estimator (SE) by using the preliminary test approach. We investigate the bias and the mean square error functions, both analytically and graphically to compare the performance of the estimators. The relative efficiency of the estimators are also studied to search for a better choice. Extensive computations have been used to produce graphs to critically check various affects on the properties of the estimators.The analysis reveals the fact that although there is no uniformly superior estimator that dominates the others, the SE dominates the other two biased estimators if the non-sample information regarding the value of $\beta_1$ is not too far from its true value. In practice, the non-sample information is usually available from past experience or expert knowledge, and hence it is expected that such an information will not be too far from the true value.

The next section deals with the specification of the model and definition of the unrestricted estimators of the slope and spread parameters as well as the derivation of the likelihood ratio test statistic. The three alternative 'improved' estimators are defined in section 3. The expressions of bias and mse functions of the estimators are obtained in section 4. Comparative study of the relative efficiency of the estimators are included in section 5. Some concluding remarks are given in section 6.

## 2 The Student-t Regression Model

Fisher (1956) discarded the normal distribution as a sole model for the distribution of errors. Fraser (1979) showed that the results based on the Student-t models for linear models are applicable to those of normal models, but not the vice-versa. Prucha and Kelejian (1984) critically analyzed the problems of normal distribution and recommended the Student-t distribution as a better alternative for many problems. The failure of the normal distribution to model the fat-tailed distributions has led to the use of the Student-t model in such a situation. In addition to being robust, the Student-t distribution is a 'more typical' member of the elliptical class of distributions. Moreover, the normal distribution is a special (limiting) case of the Student-t distribution. It also covers the Cauchy distribution on the other extreme. Extensive work on this area of non-normal models has been done in recent years. A brief summary of such literature has been given by Chmielewiski (1981), and other notable references include Fang and Zhang (1990), Khan and Haq (1990), Fang and Anderson (1990) and Celler et al (1995). Zellner (1976) first introduced the regression model with Student-t errors.

Let us express the $n$ sample responses from from a linear regression model in the following convenient form

$$\boldsymbol{y} = \beta_0 \mathbf{1}_n + \beta_1 \boldsymbol{x} + \boldsymbol{e} \tag{2.1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is an $n \times 1$ vector of responses, $\mathbf{1}_n = (1, \ldots, 1)'$ – a vector of $n$-tuple of one's, $\boldsymbol{x}$ is the $n \times 1$ vector of explanatory variable, $\beta_0$ and $\beta_1$ are the unknown intercept and slope parameters respectively and $\boldsymbol{e} = (e_1, \ldots, e_n)'$ is a vector of errors with independent components which is distributed as $N_n(\mathbf{0},\ \tau^2 I_n)$ for a given value of $\tau$. Assuming that $\tau$ follows an inverted gamma distribution with parameters $\nu$ and $\sigma^2$, the density function is given by

$$f(\tau) = \frac{2}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\sigma^2 \nu}{2}\right)^{\nu/2} (\tau)^{-(\nu+1)} e^{-\frac{\sigma^2 \nu}{2\tau^2}}, \qquad \tau > 0, \tag{2.2}$$

where $\nu$ is the shape parameter and $\sigma^2$ is the scale parameter. It is well known that the mixture distribution of the errors and $\tau$ is an $n$-dimensional Student-t distribution with shape $\nu$, location $\mathbf{0}$ and scale $\sigma^2$. We write $[\boldsymbol{e}|\tau] \sim N_n(\nu, \mathbf{0}, \tau^2)$ and $[\boldsymbol{e}] \sim t_n(\nu, \mathbf{0}, \sigma^2)$. Thus the (unconditional) density of $\boldsymbol{y}$ becomes

$$p(\boldsymbol{y}|\beta_0, \beta_1, \sigma^2) = \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{[\pi\nu\sigma^2]^{\frac{n}{2}}\, \Gamma\left(\frac{\nu}{2}\right)} \left[1 + \frac{1}{\nu\sigma^2} \sum_{j=1}^{n} (y_j - \beta_0 - \beta_1 x_j)^2\right]^{-\frac{\nu+n}{2}}. \tag{2.3}$$

Note that $\mathrm{E}[\boldsymbol{y}] = \beta_0 1_n + \beta_1 \boldsymbol{x}$ and $\mathrm{Var}[\boldsymbol{y}] = \frac{\nu}{\nu-2}\sigma^2 I_n$.

# 3 Some Preliminaries

Following Zellner (1976), from the exclusive sample information, the *unrestricted estimator* (UE) of the slope $\beta_1$ is the usual maximum likelihood estimator (mle) given by

$$\tilde{\beta}_1 = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}. \tag{3.1}$$

For the normal model, conditional on $\tau$, the sampling distribution of the mle of $\beta_1$ is normal with mean, $E(\tilde{\beta}_1) = \beta_1$ and variance, $E(\tilde{\beta}_1 - \beta_1)^2 = \frac{\tau^2}{S_{xx}}$ in which $S_{xx} = \sum_{j=1}^{n}(x_j - \bar{x})^2$. For the Student-t model $\tilde{\beta}_1$ is unbiased for $\beta_1$, and the mse is the same as its variance. Thus the bias and mse of $\tilde{\beta}_1$ are given by

$$B_1(\tilde{\beta}_1) = 0 \quad \text{ and } \quad M_1(\tilde{\beta}_1) = \frac{\nu}{\nu - 2}\frac{\sigma^2}{S_{xx}} \quad \text{ respectively.} \tag{3.2}$$

Note, unlike for the normal model, the mse of $\tilde{\beta}_1$ for the Student t model depends on the shape parameter $\nu$. We compare the above bias and mse functions with those of the three biased estimators, and search for a 'best' estimator that may perform better than the other estimators under some specific condition. It is well known that the mle of $\sigma^2$ is

$$S_n^{*2} = \frac{1}{n}(\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}), \tag{3.3}$$

where $\hat{\boldsymbol{y}} = \tilde{\beta}_0 \mathbf{1}_n + \tilde{\beta}_1 \boldsymbol{x}$ in which $\tilde{\beta}_0$ is the mle of $\beta_0$.

This estimator of $\sigma^2$ is biased. However, an unbiased estimator of $\sigma^2$ is given by

$$S_n^2 = \frac{1}{n-2}(\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}}). \tag{3.4}$$

The unbiased estimator of $\sigma^2$ has a scaled $\chi^2$ distribution with shape parameter $m = (n-2)$.

To be able to use the uncertain non-sample prior information in the estimation of the slope, it is essential to remove the element of uncertainty concerning it's value. Fisher suggested to express the uncertain non-sample prior information in the form of a null hypothesis, $H_0 : \beta_1 = \beta_{10}$ and treat it as a nuisance parameter. He proposed to conduct an appropriate statistical test on the null-hypothesis against the alternative $H_A : \beta_1 \neq \beta_{10}$ to remove the uncertainty in the non-sample prior information. For the problem under study, an appropriate test is the likelihood ratio test (LRT). The LRT for testing the null-hypothesis is given by the test statistic

$$\mathcal{L}^2 = \frac{S_{xx}(\tilde{\beta}_1 - \beta_{10})^2}{S_n^2}. \tag{3.5}$$

Since $\frac{\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})}{S_n}$ follows a Student-t distribution under $H_0$ and non-central Student-t distribution under $H_A$, the above statistic $\mathcal{L}^2$, under $H_A$, follows a non-central $F$-distribution with $(1, m)$ degrees of freedom (d.f.) in which $m = n-2$, and non-centrality parameter $\frac{1}{2}\Delta^2$, where

$$\Delta^2 = \frac{S_{xx}(\beta_1 - \beta_{10})^2}{\sigma^2} = \frac{\delta^2}{\sigma^2} \tag{3.6}$$

where $\delta = \sqrt{S_{xx}}|(\beta_1 - \beta_{10})|$. This test statistic would be used to define the PTE, and the shrinkage estimator by following the preliminary test approach to the shrinkage estimation.

# 4 Alternative Estimators of the Slope

In this section we use the uncertain non-sample prior information and the coefficient of distrust on the null hypothesis to estimate the slope parameter. First we combine the exclusive sample based estimator, $\tilde{\beta}_1$ with the non-sample prior information presented in the form of a null hypothesis, $H_0 : \beta_1 = \beta_{10}$ in some reasonable way. Now, consider a simple linear combination of $\beta_{10}$ and $\tilde{\beta}_1$ as

$$\hat{\beta}_1(d) = d\tilde{\beta}_1 + (1-d)\beta_{10}, \quad 0 \leq d \leq 1. \tag{4.1}$$

This estimator of $\beta_1$ is called the *restricted estimator* (RE), where $d$ is the *degree of distrust* in the null hypothesis, $H_0 : \beta_1 = \beta_{10}$. Here, $d = 0$, means there is *no distrust* in the $H_0$ and we get $\hat{\beta}_1(d = 0) = \beta_{10}$, while $d = 1$ means there is *complete distrust* in the $H_0$ and we get $\hat{\beta}_1(d=1)=\tilde{\beta}_1$. If $0<d<1$, the degree of distrust is an intermediate value which results in an interpolated value between $\beta_{10}$ and $\tilde{\beta}_1$ given by (3.1).

Following Bancroft (1944) we define the shrinkage preliminary test estimator (SPTE) of the slope parameter as

$$\begin{aligned}\hat{\beta}_1^{\mathrm{PTE}}(d) &= \hat{\beta}_1 I(F<F_\alpha)+\tilde{\beta}_1 I(F\geq F_\alpha)\\ &= \tilde{\beta}_1-(1-d)(\tilde{\beta}_1-\beta_{10})I(F<F_\alpha), \end{aligned} \quad (4.2)$$

where $I(A)$ is an indicator function of the set $A$ and $F_\alpha$ is the critical value chosen for the $\alpha$-level test based on the $F$-distribution with $(1,m)$ degrees of freedom. A simplified form of the above SPTE is the preliminary test estimator (PTE)

$$\hat{\beta}_1^{\mathrm{PTE}}=\beta_{10}I(F<F_\alpha)+\tilde{\beta}_1 I(F\geq F_\alpha), \quad (4.3)$$

which is a special case of (4.2) when $d=0$. Note that, the $\hat{\beta}_1^{\mathrm{PTE}}(d)$ is a convex combination of $\hat{\beta}_1(d)$ and $\tilde{\beta}_1$, and $\hat{\beta}_1^{\mathrm{PTE}}(d=0)$ is a convex combination of $\beta_{10}$ and $\tilde{\beta}_1$. We may rewrite (4.3) as

$$\hat{\beta}_1^{\mathrm{PTE}}(d)=\tilde{\beta}_1-(1-d)(\tilde{\beta}_1-\beta_{10})I(F<F_\alpha), \quad (4.4)$$

where $F_\alpha$ is the $(1-\alpha)^{th}$ quantile of a central $F$-distribution with $(1,\ m)$ degrees of freedom. For $d=0$, we get (4.3) as

$$\hat{\beta}_1^{\mathrm{PTE}}(d=0)=\tilde{\beta}_1-(\tilde{\beta}_1-\beta_{10})I(F<F_\alpha). \quad (4.5)$$

The PTE is an extreme choice between $\hat{\beta}_1(d)$ and $\tilde{\beta}_1$. Hence it does not allow any smooth transition between it's two extreme values. Also, it depends on the pre-selected level of significance, $\alpha$ of the test. To overcome these problems, we consider the shrinkage estimator (SE) of $\beta_1$ defined as follows:

$$\hat{\beta}_1^{\mathrm{S}}=\beta_{10}+\left\{1-\frac{\mathrm{c}S_{\mathrm{n}}^2}{S_{\mathrm{xx}}(\tilde{\beta}_1-\beta_{10})^2}\right\}(\tilde{\beta}_1-\beta_{10}). \quad (4.6)$$

Note that in this estimator $c$ is a constant function of $n$. Now, if $F=\left\{\frac{S_{xx}(\tilde{\beta}_1-\beta_{10})^2}{S_n^2}\right\}$ is large, $\hat{\beta}_1^{\mathrm{S}}$ tends towards $\tilde{\beta}_1$, while for small $F$ equaling $c$, $\hat{\beta}_1^{\mathrm{S}}$ tends towards $\beta_{10}$ similar to the preliminary test estimator. The shrinkage estimator does not depend on the level of significance, unlike the preliminary test estimator.

# 5 Some Statistical Properties

The bias and the mean square error (mse) functions of the SE and SPTE are derived here. Also, we discuss some of the important features of these functions.

### 5.1 The Bias and MSE of RE

First the bias and the mse of the RE, $\hat{\beta}_1(d)$ are found to be

$$B_2[\hat{\beta}_1(d)] = -\frac{\sigma}{\sqrt{S_{xx}}}(1-d)\Delta, \text{ with } \Delta = \frac{\sqrt{S_{xx}}(\beta_1 - \beta_{10})}{\sigma} \tag{5.1}$$

$$M_2[\hat{\beta}_1(d)] = \frac{\sigma_*^2}{S_{xx}}[d^2 + (1-d)^2\Delta^*] \text{ with } \sigma_*^2 = \frac{\nu}{\nu-2}\sigma^2, \tag{5.2}$$

where $\Delta^* = \frac{\nu-2}{\nu}\Delta^2$ is the *departure constant* from the null-hypothesis. The value of this constant is 0 when the null hypothesis is true; otherwise it is always positive. The statistical properties of the three biased estimators depend on the value of the above departure constant. The performance of the estimators change with the change in the value of $\Delta$. We investigate this feature in a greater detail in the forthcoming sections.

### 5.2 The Bias and MSE of the SPTE

The bias function of the SPTE is given by the following theorem.

**Theorem 5.1:** *For the simple regression model with Student-t errors the bias function of the SPTE of the slope parameter is given by*

$$B_3[\hat{\beta}_1^{\text{PTE}}(d)] = -(1-d)\frac{\sigma\Delta}{\sqrt{S_{xx}}}G^*_{3,m}\Big(\frac{1}{3}F_\alpha; \Delta^*\Big), \tag{5.3}$$

*where*

$$G^*_{a,m}(l_\alpha; \Delta^*) = \sum_{r=0}^{\infty} I^1_{h_\alpha}\left(\frac{m}{2}; \frac{a+2r}{2}\right)\xi_r(\nu)\xi_r(\nu, \Delta^*) \tag{5.4}$$

*in which* $l_\alpha = \dfrac{a}{a+2}F_{a,m}(\alpha)$ *with* $F_{a,m}(\alpha)$ *being the* $(1-\alpha)$*-th quantile of a central* $F$*-distribution with* $a$ *and* $m$ *d.f.;* $h_\alpha = \dfrac{am}{am + (a+2r)F_{a,m}(\alpha)}$ *for* $a = 1$*; and* $I^1_{h_\alpha}\left(\frac{m}{2}; \frac{a+2r}{2}\right) = 1 - I_0^{h_\alpha}\left(\frac{m}{2}; \frac{a+2r}{2}\right)$ *is the incomplete beta function evaluated at* $h_\alpha$*;*

$$\xi_r(\nu) = \frac{\Gamma\left(\frac{\nu}{2}+r\right)}{r!\Gamma\left(\frac{\nu}{2}\right)}; \ \xi_r(\nu, \Delta^*) = \frac{(\Delta^*/\nu-2)^r}{[1+\Delta^*/(\nu-2)]^{\nu/2+r}} \text{ with } \Delta^* = \frac{\nu-2}{\nu}\Delta^2. \tag{5.5}$$

**Proof:** From the definition, for given $\tau$, the expression of bias of the SPTE is

$$\begin{aligned} E[(\hat{\beta}_1^{\text{PTE}}(d) - \beta_1)|\tau] &= E(\tilde{\beta}_1 - \beta_1) - (1-d)E\{(\tilde{\beta}_1 - \beta_{10})I(F < F_\alpha)\} \qquad (5.6)\\ &= -(1-d)\frac{\tau}{\sqrt{S_{xx}}}E\Big\{\frac{\sqrt{S_{xx}}(\tilde{\beta}_1-\beta_{10})}{\tau}I\Big(\frac{S_{xx}(\tilde{\beta}_1-\beta_{10})^2}{S_n^2} < F_\alpha\Big)\Big\}. \end{aligned}$$

Now, conditional on $\tau$, $Z = \sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})/\tau$ is distributed as $N(\Delta_\tau, 1)$, where $\Delta_\tau = \frac{\sqrt{S_{xx}}}{\tau}(\beta_1 - \beta_{10})$, and $S_{xx}(n-2)S_n^2/\tau^2$ is distributed (independently) as a central chi-square variable with $m$ degrees of freedom. To evaluate the above expression we need the following lemma.

**Lemma 5.1**. *If $Z \sim \mathcal{N}(\Lambda, 1)$ and $\phi(Z^2)$ is a Borel measurable function, then*

$$E\{Z\phi(Z^2)\} = \Lambda E\phi[\chi_3^2(\Lambda^2)]. \tag{5.7}$$

The proof of the lemma can be found in Saleh (2006) or Appendix B2 of Judge and Bock (1978).

Evaluating the expression in (5.6), conditional on $\tau$, the bias function of $\hat{\beta}_1^{\text{PTE}}(d)$ is found to be

$$B_3[(\hat{\beta}_1^{\text{PTE}}(d))|\tau] = -(1-d)\frac{\tau}{\sqrt{S_{xx}}}\Delta_\tau G_{3,m}\Big(\frac{1}{3}F_\alpha; \Delta_\tau^2\Big), \tag{5.8}$$

where conditional on $\tau$, $G_{3,m}(\cdot; \Delta_\tau^2)$ is the c.d.f. of a non-central F-distribution with $(3, m)$ degrees of freedom and non-centrality parameter $\frac{1}{2}\Delta_\tau^2 = \frac{\delta^2}{2\tau^2}$. The above c.d.f. involves incomplete beta function with appropriate arguments. For the Student-t model the above expression for the cdf is not valid. The appropriate expression for the Student-t model has been given in (5.4). The proof of the theorem is completed by taking expectation of $B_3[(\hat{\beta}_1^{\text{PTE}}(d))|\tau]$ with respect to $\tau$. The bias function of the SPTE depends on the *coefficient of distrust* and the *departure constant*, among other things.

**Theorem 5.2:** *For the Student-t regression model the mse of the SPTE is given by*

$$\begin{aligned} M_3[\hat{\beta}_1^{\text{PTE}}(d)] = \frac{\sigma_*^2}{S_{xx}}\Big[&1 - (1-d^2)G_{3,m}^*\Big(\frac{1}{3}F_\alpha; \Delta^*\Big) \\ &+(1-d)\Delta^*\Big\{2G_{3,m}^*\Big(\frac{1}{3}F_\alpha; \Delta^*\Big) - (1+d)G_{5,m}^*\Big(\frac{1}{5}F_\alpha; \Delta^*\Big)\Big\}\Big], \end{aligned} \tag{5.9}$$

*where $\sigma_*^2 = \frac{\nu-2}{\nu}\sigma^2$ and $G_{a,m}^*\Big(\frac{1}{a}F_\alpha; \Delta^*\Big)$ has been defined in (5.4).*

**Proof:** From the definition, conditional on $\tau$, the mse expression of the SPTE is given by

$$\begin{aligned} M_3\Big[(\hat{\beta}_1^{\text{PTE}}(d))|\tau\Big] &= E\Big[\hat{\beta}_1^{\text{PTE}}(d) - \beta_1\Big]^2 \\ &= E(\tilde{\beta}_1 - \beta_1)^2 + (1-d)^2 E(\tilde{\beta}_1 - \beta_{10})^2 I(F < F_\alpha) \\ &\quad -2(1-d)E[(\tilde{\beta}_1 - \beta_1)(\beta_1 - \beta_{10})]I(F < F_\alpha) \\ &= \frac{\tau^2}{S_{xx}} + (1-d)^2 E[(\tilde{\beta}_1 - \beta_{10})^2 I(F < F_\alpha)] \\ &\quad -2(1-d)E\Big[\{(\tilde{\beta}_1 - \beta_{10}) - (\beta_1 - \beta_{10})\}(\tilde{\beta}_1 - \beta_{10})I(F < F_\alpha)\Big]. \end{aligned} \tag{5.10}$$

To evaluate the above expression of the mean square error of $\hat{\beta}_1^{\text{PTE}}(d)$ we need the following lemma.

**Lemma 5.2**. *If $Z \sim \mathcal{N}(\Lambda, 1)$ and $\phi(Z^2)$ is a Borel measurable function, then*

$$E[Z^2\phi(Z^2)] = E\Big[\phi\{\chi_3^2(\Lambda^2)\}\Big] + \Lambda^2 E\Big[\phi\{\chi_5^2(\Lambda^2)\}\Big]. \tag{5.11}$$

The proof of the lemma is given in Saleh (2006) or Appendix B2 of Judge and Bock (1978).

After completing the evaluation of all the terms on the R.H.S. of the expression of the mse function of the SPTE, we get

$$M_3[(\hat{\beta}_1^{\mathrm{PTE}}(d))|\tau] = \frac{\tau^2}{S_{xx}}\Big[1-(1-d^2)G_{3,m}\Big(\frac{1}{3}F_\alpha;\Delta_\tau^2\Big) \tag{5.12}$$
$$+(1-d)\Delta_\tau^2\Big\{2G_{3,m}\Big(\frac{1}{3}F_\alpha;\Delta_\tau^2\Big)-(1+d)G_{5,m}\Big(\frac{1}{5}F_\alpha;\Delta_\tau^2\Big)\Big\}\Big].$$

The proof of the theorem is completed by taking expectation on $M_3[(\hat{\beta}_1^{\mathrm{PTE}}(d))|\tau]$ with respect to $\tau$.

Figure 1, displays the behavior of the mse based relative efficiency functions of the SE and SPTE for a fixed $\alpha$ with the change in the value of $\Delta^2$. The four graphs illustrate the different features of the relative efficiency functions for selected values of the *coefficient of distrust*, $d = 0.00, 0.25, 0.50, 1.00$ when $\nu$ is fixed at 5.

Figure 2, displays the behavior of the mse based relative efficiency functions of the SE and SPTE for a fixed $\alpha$ with the change in the value of $\Delta^2$. The four graphs illustrate the different features of the relative efficiency functions for selected values of the *degrees of freedom*, $\nu = 5,\ 10, 20,\ 40$ when $d$ is fixed at 0.50.

**Some Properties of the MSE of SPTE**

(a) Under the null hypothesis $\Delta^2 = 0$, and hence the mse of $\hat{\beta}_1^{\mathrm{PTE}}(d)$ equals

$$\frac{\sigma_*^2}{S_{xx}}\left[1-(1-d^2)G_{3,m}^*\left(\frac{1}{3}F_\alpha;0\right)\right] < \frac{\sigma_*^2}{S_{xx}}. \tag{5.13}$$

Thus, at $\Delta^2 = 0$ SPTE of $\beta_1$ performs better than $\tilde{\beta}_1$, the UE. As $\alpha \to 0$, $G_{3,m}^*\Big(\frac{1}{3}F_\alpha;0\Big) \to 1$, then

$$\frac{\sigma_*^2}{S_{xx}}\left[1-(1-d^2)G_{3,m}^*\left(\frac{1}{3}F_\alpha;0\right)\right] \to \frac{d^2\sigma_*^2}{S_{xx}}, \tag{5.14}$$

which is the mse of $\hat{\beta}_1(d)$. On the other hand, if $F_\alpha \to 0$, $G_{3,m}^*\Big(\frac{1}{3}F_\alpha;0\Big) \to 0$, then

$$\frac{\sigma_*^2}{S_{xx}}\left[1-(1-d^2)G_{3,m}^*\left(\frac{1}{3}F_\alpha;0\right)\right] \to \frac{\sigma_*^2}{S_{xx}}, \quad \text{which is the mse of } \tilde{\beta}_1. \tag{5.15}$$

(b) As $\Delta^2 \to \infty$, $G_{a,m}^*\Big(\frac{1}{a}F_\alpha;\Delta^*\Big) \to 0$, (for $a = 3,\ 5$) this means the expression at (5.9) tends towards $\frac{\sigma_*^2}{S_{xx}}$, the mse of the UE.

(c) Since $G_{3,m}^*\Big(\frac{1}{3}F_\alpha;\Delta^*\Big)$ is always greater than $G_{5,m}^*\Big(\frac{1}{5}F_\alpha;\Delta^*\Big)$ for any value of $\alpha$, replacing $G_{5,m}^*\Big(\frac{1}{5}F_\alpha;\Delta^*\Big)$ by $G_{3,m}^*\Big(\frac{1}{3}F_\alpha;\Delta^2\Big)$, (5.4) becomes

$$\geq \frac{\sigma_*^2}{S_{xx}}\Big[1+(1-d^2)G_{3,m}^*\Big(\frac{1}{3}F_\alpha;\Delta^*\Big)\{(1-d)\Delta^2-(1+d)\}\Big] \tag{5.16}$$
$$\geq \frac{\sigma_*^2}{S_{xx}} \quad \text{whenever} \quad \Delta^2 > \frac{1+d}{1-d}.$$

On the other hand, (5.9) may be rewritten as

$$\frac{\sigma_*^2}{S_{xx}}\Big[1+(1-d)G_{3,m}^*\Big(\frac{1}{3}F_\alpha;\Delta^*\Big)\{2\Delta^2-(1+d)\}-(1-d^2)G_{5,m}^*\Big(\frac{1}{5}F_\alpha;\Delta^*\Big)\Big] \quad (5.17)$$

$$\leq \frac{\sigma_*^2}{S_{xx}} \text{ whenever } \Delta^2 < \frac{1+d}{2}. \quad (5.18)$$

This means that the mse of $\hat{\beta}_1^{\mathrm{PTE}}(d)$ as a function of $\Delta^2$ crosses the constant line $M_1(\tilde{\beta}_1) = \frac{\sigma_*^2}{S_{xx}}$ in the interval $\left(\frac{1+d}{2}, \frac{1+d}{1-d}\right)$.

(d) A general picture of the mse graph may be described as follows: The mse-function begins with the smallest value $\frac{\sigma_*^2}{S_{xx}}\Big[1-(1-d^2)G_{3,m}^*\Big(\frac{1}{3}F_\alpha;0\Big)\Big]$ at $\Delta^2=0$. As $\Delta^2$ grows large, the function increases monotonically crossing the constant line $\frac{\sigma_*^2}{S_{xx}}$ in the interval $\left(\frac{1+d}{2}, \frac{1+d}{1-d}\right)$ and reaches a maximum in the interval $\left(\frac{1+d}{1-d}, \infty\right)$ then monotonically decreases towards $\frac{\sigma_*^2}{S_{xx}}$ as $\Delta^2 \to \infty$.

## 5.3 The Bias and MSE of SE

Following Balforine and Zacks (1992) we compute the bias and the mse of the SE, $\hat{\beta}_1^{\mathrm{S}}$.

**Theorem 5.3:** *For the Student-t regression model the bias of the SE of the slope is given by*

$$B_4(\hat{\beta}_1^{\mathrm{S}}) = \frac{-\sigma}{\sqrt{S_{xx}}}cK_n\{2\Phi(\Delta)-1\}, \quad (5.19)$$

*where* $K_n = \sqrt{\frac{2}{n-2}}\,\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})}$.

**Proof:** Conditional on $\tau$, the bias of the SE is defined by

$$\begin{aligned} E[\hat{\beta}_1^{\mathrm{S}}-\beta_1] &= -cE\left[\frac{S_n(\tilde{\beta}_1-\beta_{10})}{|\sqrt{S_{xx}}(\tilde{\beta}_1-\beta_{10})|}\right] \\ &= -\frac{c}{\sqrt{S_{xx}}}E[S_n]E\left\{\frac{Z}{|Z|}\right\}, \end{aligned} \quad (5.20)$$

where $Z = \frac{\sqrt{S_{xx}}(\tilde{\beta}_1-\beta_{10})}{\tau} \sim \mathcal{N}(\Delta_\tau, 1)$. We use the following lemma to evaluate $E\left\{\frac{Z}{|Z|}\right\}$.

**Lemma 5.3**. *If $Z \sim \mathcal{N}(\Lambda, 1)$ and $\phi(Z^2)$ is a Borel measurable function, then*

$$E\left\{\frac{Z}{|Z|}\right\} = 1-2\Phi(-\Lambda) \quad (5.21)$$

*where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution.* The proof of the lemma is straightforward. Note that conditional on $\tau$, $\frac{mS_n^2}{\tau^2} \sim \chi_m^2$ and hence $E[S_n] = \sqrt{\frac{2}{n-2}}\frac{\Gamma(n-1)}{\Gamma(\frac{n-2}{2})}\tau$.

So, for a given $\tau$, the bias of the SE becomes

$$B_4(\hat{\beta}_1^{\mathrm{S}}|\tau) = -\frac{c}{\sqrt{S_{xx}}}K_n\tau\{2\Phi(\Delta_\tau)-1\}. \quad (5.22)$$

Expectation of $B_4(\hat{\beta}_1^{\mathrm{S}}|\tau)$ with respect to $\tau$ completes the proof. From the expression of the bias function, the quadratic bias of the SE, $QB_4(\hat{\beta}_1^{\mathrm{S}})$ is obtained as

$$QB_4(\hat{\beta}_1^{\mathrm{S}}) = \frac{\sigma^2}{S_{xx}} c^2 K_n^2 \{2\Phi(\Delta) - 1\}^2. \tag{5.23}$$

As $\Delta^2 \to 0$, $QB_4(\hat{\beta}_1^{\mathrm{S}}) \to 0$ and as $\Delta^2 \to \infty$, $QB_4(\hat{\beta}_1{}^{\mathrm{S}}) \to \frac{\sigma_*^2}{S_{xx}} K_n^2 c^2$. Therefore, $QB_4(\hat{\beta}_1^{\mathrm{S}})$ is a non-decreasing monotonic function of $\Delta^2$. Thus, unless $\Delta^2$ is near the origin, the quadratic bias of the SE is significantly large.

**Theorem 5.4:** *For the Student-t regression model the mse of the SE of the slope is given by*

$$M_4(\hat{\beta}_1^{\mathrm{S}}) = \frac{\sigma_*^2}{S_{xx}} \Big\{ 1 - \frac{2}{\pi} K_n^2 \Big[ 2\eta(\Delta^*) - 1 \Big] \Big\}, \tag{5.24}$$

*where* $\eta(\Delta^*) = \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{\Delta^*}{\nu-2}\right)^{-\frac{\nu}{2}}$

**Proof:** From the definition, conditional on $\tau$, the mse of $\hat{\beta}_1^{\mathrm{S}}$ is

$$\begin{aligned} E[(\hat{\beta}_1^{\mathrm{S}} - \beta_1)^2|\tau] &= E(\tilde{\beta}_1 - \beta_1)^2 + c^2 E(S_n^2) E\left\{ \frac{(\tilde{\beta}_1 - \beta_{10})^2}{[\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})]^2} \right\} \\ &\quad -2cE\Big\{ \frac{(\tilde{\beta}_1 - \beta_1)(\tilde{\beta}_1 - \beta_{10})}{|\sqrt{S_{xx}}(\tilde{\beta}_1 - \beta_{10})|} \Big\} E(S_n) \\ &= \frac{\tau^2}{S_{xx}} + \frac{c^2\tau^2}{S_{xx}} - 2c\frac{\tau^2 K_n}{S_{xx}} \Big\{ E(|Z|) - \Delta_\tau E\Big(\frac{Z}{|Z|}\Big) \Big\}, \end{aligned} \tag{5.25}$$

where $Z \sim \mathcal{N}(\Delta_\tau, 1)$. To find $E(|Z|)$, we have the following lemma.

**Lemma 5.4**. *If* $Z \sim \mathcal{N}(\Lambda, 1)$, *then*

$$E(|Z|) = \sqrt{\frac{2}{\pi}} e^{-\Lambda^2/2} + \Lambda\{2\Phi(\Lambda) - 1\} \tag{5.26}$$

*where* $\Phi(\cdot)$ *is the c.d.f. of the standard normal variable.* See Khan and Saleh (2001) for the proof of the above theorem.

Therefore, the mse of $\hat{\beta}_1^{\mathrm{S}}$ is given by

$$M_4(\hat{\beta}_1{}^{\mathrm{S}}) = \frac{\tau^2}{S_{xx}} \Big\{ 1 + c^2 - 2cK_n \sqrt{\frac{2}{\pi}} e^{-\Delta_\tau^2/2} \Big\}. \tag{5.27}$$

The value of $c$ which minimizes (5.27) depends on $\Delta_\tau^2$ and is given by

$$c^* = \sqrt{\frac{2}{\pi}} K_n e^{-\Delta_\tau^2/2}. \tag{5.28}$$

To make $c^*$ independent of $\Delta_\tau^2$, we choose $c^0 = \sqrt{\frac{2}{\pi}} K_n$. Thus, optimum $M_4(\hat{\beta}_1^{\mathrm{S}})$ reduces to

$$M_4(\hat{\beta}_1^{\mathrm{S}}) = \frac{\tau^2}{S_{xx}} \Big\{ 1 - \frac{2}{\pi} K_n^2 \Big[ 2e^{-\Delta_\tau^2/2} - 1 \Big] \Big\}. \tag{5.29}$$

Expectation of the above expression with respect to $\tau$ completes the proof. Note that $\eta(\Delta^*) = \mathrm{E}_\tau \left[ e^{-\frac{\Delta_\tau^2}{2}} \right] = \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu)} \left(1 + \frac{\Delta^*}{\nu-2}\right)^{-\frac{\nu}{2}}$.

# 6 Comparative Study

In this Section we compare the bias of the three estimators. Also, we define the relative efficiency functions of the estimators, and analyze these functions to compare the relative performances of the estimators.

## 6.1 Comparing Quadratic Bias Functions

First, we note that the quadratic bias of the RE, SPTE and SE are given by

$$\begin{aligned} QB_2[\hat{\beta}_1(d)] &= \frac{\sigma_*^2}{S_{xx}}(1-d)^2\Delta^2 \\ QB_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] &= \frac{\sigma_*^2}{S_{xx}}(1-d)^2\Delta^2\Big\{G_{3,m}^*\Big(\frac{1}{3}F_\alpha;\Delta^*\Big)\Big\}^2 \\ QB_4[\hat{\beta}_1^{\mathrm{S}}] &= \frac{\sigma_*^2}{S_{xx}}c^2K_n^2\{2\Phi(\Delta)-1\}^2. \end{aligned} \tag{6.1}$$

Clearly, under the null-hypothesis $QB_2[\hat{\beta}_1(d)] = QB_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] = QB_4[\hat{\beta}_1^{\mathrm{S}}(d)] = 0$ for all $d$ and $\alpha$.

When $\Delta \to \infty$, $QB_2[\hat{\beta}_1(d)] \to \infty$ except at $d = 1$; $QB_3[\hat{\beta}_1^{\mathrm{PTE}}(d)] \to 0$ for all $\alpha$ and $d$; and $QB_4[\hat{\beta}_1{}^{\mathrm{S}}] \to \frac{\sigma_*^2}{S_{xx}}c^2K_n^2$, a constant that does not depend on $d$. Therefore, in terms of quadratic bias, RE is uniformly dominated by both the SPTE and SE. For very large values of $\Delta$, the SE is dominated by the SPTE regardless of the value of $\alpha$. From small to moderate values of $\Delta$, there is no uniform domination of one estimator over the other. In this case, domination depends on the level of significance, $\alpha$. For small values of $\alpha$, the SPTE is dominated by the SE and for larger values of $\alpha$, the SE is dominated by the SPTE. However, Chiou and Saleh (2002) suggest the value of $\alpha$ to be between 20% and 25%. In this interval of $\alpha$, the quadratic bias of the SPTE approaches to zero for not too small values of $\Delta$. However, in practice, the non-centrality parameter is unlikely to be very large (otherwise the credibility of prior information is in serious question) and $\alpha$ is usually preferred to be small. The quadratic bias of the SE is relatively stable and approaches to a constant value starting from some moderate value of $\Delta$ and is unaffected by the choice of $d$ and $\alpha$. Therefore, the SE may be a better choice among the biased estimators considered in this paper.

## 6.2 The Relative Efficiency

First we define the relative efficiency functions of the biased estimators as the ratio of the reciprocal of the mse functions. Then we compare the relative performance of the estimators by using the relative efficiency criterion.

**Comparing RE against UE**

The relative efficiency of $\hat{\beta}_1(d)$ compared to $\tilde{\beta}_1$ is denoted by $\text{RE}[\hat{\beta}_1(d) : \tilde{\beta}_1]$ and is obtained as

$$\text{RE}[\hat{\beta}_1(d) : \tilde{\beta}_1] = [d^2 + (1-d)^2\Delta^2]^{-1}. \tag{6.2}$$

We observe the following based on (6.2).

(i) If the non-sampling information is correct, i.e., $\Delta^2 = 0$, the $\text{RE}[\hat{\beta}_1(d) : \tilde{\beta}_1] = d^{-2} > 1$ and $\hat{\beta}_1(d)$ is more efficient than $\tilde{\beta}_1$. Thus, under the null hypothesis the biased estimator, RE performs better than the unbiased estimator, UE.

(ii) If the non-sampling information is incorrect, i.e., $\Delta^2 > 0$ we study the expression in (6.2) as a function of $\Delta^2$ for a fixed $d$-value. As a function of $\Delta^2$, (6.2) is a decreasing function with its maximum value $d^{-2}(> 1)$ at $\Delta^2 = 0$ and minimum value 0 at $\Delta^2 = +\infty$. It equals 1 at $\Delta^2 = \frac{1+d}{1-d}$. Thus, if $\Delta^2 \in [0, \frac{1+d}{1-d})$, $\hat{\beta}_1(d)$ is more efficient than $\tilde{\beta}_1$, and outside this interval $\tilde{\beta}_1$ is more efficient than $\hat{\beta}_1(d)$. For example, if $d = \frac{1}{2}$, the interval in which $\hat{\beta}_1(d)$ is more efficient than $\tilde{\beta}_1$ is $[0, 3)$, while $\tilde{\beta}_1$ is more efficient in $[3, \infty)$ than $\hat{\beta}_1(d)$. For d$= 0.5$ the maximum efficiency of $\hat{\beta}_1(d)$ over $\tilde{\beta}_1$ is 4.

**Comparing SPTE against UE**

Now, we consider the relative efficiency of the SPTE compared to the UE. It is given by

$$\begin{aligned} RE[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] &= \Big[1 - (1-d^2)G^*_{3,m}\Big(\frac{1}{3}F_\alpha; \Delta^*\Big) + (1-d)\Delta^2 \\ &\quad \times\Big\{2G^*_{3,m}\Big(\frac{1}{3}F_\alpha; \Delta^*\Big) - (1+d)G^*_{5,m}\Big(\frac{1}{5}F_\alpha; \Delta^*\Big)\Big\}\Big]^{-1} \end{aligned} \tag{6.3}$$

for any fixed $d$ $(0 \le d \le 1)$ and at a fixed level of significance $\alpha$. As $F_\alpha \to \infty$, $RE[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] \to [1 - (1-d^2) + (1-d)^2\Delta^2]^{-1} = [d^2 + (1-d)^2\Delta^2]^{-1}$ which is the relative efficiency of $\hat{\beta}_1(d)$ compared to $\tilde{\beta}_1$. On the other hand, as $F_\alpha \to 0$, $RE[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] \to 1$. This means the relative efficiency of the SPTE is the same as the unrestricted estimator, $\tilde{\beta}_1$. Note that under the null hypothesis, $\Delta^2 = 0$, the relative efficiency expression (6.3) equals

$$\Big[1 - (1-d^2)G^*_{3,m}\Big(\frac{1}{3}F_\alpha; 0\Big)\Big]^{-1} \ge 1, \tag{6.4}$$

which is the maximum value of the relative efficiency. Thus the relative efficiency function monotonically decreases crossing the 1-line for $\Delta^2$-value between $\frac{1+d}{2}$ and $\frac{1+d}{1-d}$, to a minimum for some $\Delta^2 = \Delta^2_{\text{min}}$ and then monotonically increases, to approach the unit value from below. The relative efficiency of the preliminary test estimator equals unity whenever the value of $\Delta^2$ is

$$\Delta^2_* = \frac{(1+d)}{\Big\{2 - (1+d)\frac{G^*_{5,m}(\frac{1}{5}F_\alpha;\Delta^*)}{G^*_{3,m}(\frac{1}{3}F_\alpha;\Delta^*)}\Big\}}, \tag{6.5}$$

where $\Delta_*^2$ lies in the interval $\left(\frac{1+d}{2}, \frac{1+d}{1-d}\right)$. This means that

$$RE\Big[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1\Big] \lesseqgtr 1 \text{ according as } \Delta_*^2 \lesseqgtr \Delta^2. \tag{6.6}$$

Finally, as $\Delta^2 \to \infty$, $\text{RE}[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1] \to 1$. Thus, the preliminary test estimator is more efficient than the unrestricted estimator whenever $\Delta^2 < \Delta_*^2$, otherwise $\tilde{\beta}_1$ is more efficient than SPTE up to a moderate value of $\Delta^2$. As for the relative efficiency of $\hat{\beta}_1^{\text{PTE}}(d)$ compared to $\hat{\beta}_1(d)$ we have

$$\text{RE}\Big[\hat{\beta}_1^{\text{PTE}}(d) : \hat{\beta}_1\Big] = [d^2 + (1-d)^2\Delta^2][1 + g(\Delta^2)]^{-1}, \tag{6.7}$$

where

$$\begin{aligned} g(\Delta^2) &= (1-d)\Delta^2\Big\{2G_{3,m}^*\Big(\tfrac{1}{3}F_\alpha; \Delta^*\Big) - (1+d)G_{5,m}^*\Big(\tfrac{1}{5}F_\alpha; \Delta^*\Big)\Big\} \\ &\qquad -(1+d^2)G_{3,m}^*\Big(\tfrac{1}{3}F_\alpha; \Delta^*\Big). \end{aligned} \tag{6.8}$$

Under the null-hypothesis,

$$\text{RE}[\hat{\beta}_1^{\text{PTE}}(d) : \hat{\beta}_1(d)] = d^2\Big[1 - (1-d^2)G_{3,m}^*\Big(\frac{1}{3}F_\alpha; 0\Big)\Big]^{-1} \geq d^2. \tag{6.9}$$

At the same time we consider the result at (6.3). In combination, we obtain

$$d^2 \leq \text{RE}[\hat{\beta}_1^{\text{PTE}}(d) : \hat{\beta}_1(d)] \leq 1 \leq \text{RE}[\hat{\beta}_1^{\text{PTE}}(d) : \tilde{\beta}_1]. \tag{6.10}$$

For general $\Delta^2 > 0$, we have

$$\text{RE}[\hat{\beta}_1^{\text{PTE}}(d) : \hat{\beta}_1(d)] \lesseqgtr 1 \text{ according as} \tag{6.11}$$

$$\Delta^2 \lesseqgtr \frac{1+d}{1-d} \; \frac{\Big\{1 - G_{3,m}^*\Big(\frac{1}{3}F_\alpha; \Delta^*\Big)\Big\}}{\Big\{1 - 2G_{3,m}^*\Big(\frac{1}{3}F_\alpha; \Delta^*\Big) - (1+d)G_{5,m}^*\Big(\frac{1}{5}F_\alpha; \Delta^*\Big)\Big\}}. \tag{6.12}$$

Finally, as $\Delta^2 \to \infty$, $\text{RE}[\hat{\beta}_1^{\text{PTE}}(d); \hat{\beta}_1(d)] \to 0$. Thus, except for a small interval around 0, $\hat{\beta}_1^{\text{PTE}}(d)$ is more efficient than $\hat{\beta}_1(d)$.

**Comparing SE against UE**

The relative efficiency of $\hat{\beta}_1^{\text{S}}$ compared to $\tilde{\beta}_1$ is given by

$$\text{RE}(\hat{\beta}_1^{\text{S}} : \tilde{\beta}_1) = \Big[1 - \frac{2}{\pi}\text{K}_\text{n}^2\Big\{2\text{e}^{-\Delta^2/2} - 1\Big\}\Big]^{-1}. \tag{6.13}$$

Under the null-hypothesis $\Delta^2 = 0$, and hence

$$\text{RE}(\hat{\beta}_1^{\text{S}} : \tilde{\beta}_1) = \Big[1 - \frac{2}{\pi}\text{K}_\text{n}^2\Big]^{-1} \geq 1. \tag{6.14}$$

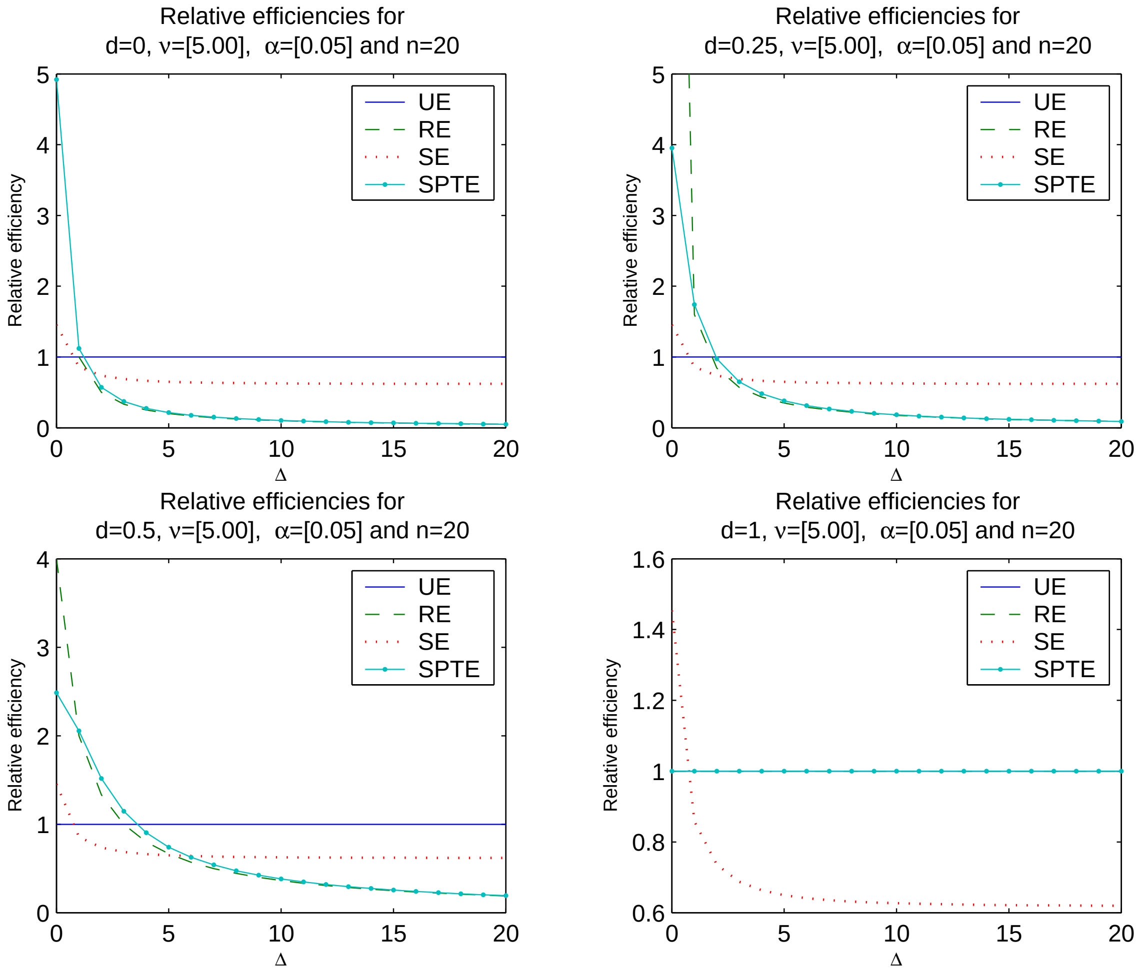


Figure 1: Graph of relative efficiency of RE, SPTE and SE relative to UE for $\nu = 5$ and different $d$

In general, $\text{RE}(\hat{\beta}_1^{\text{S}} : \tilde{\beta}_1)$ decreases from $[1 - \frac{2}{\pi}K_n^2]^{-1}$ at $\Delta^2 = 0$ and crosses the 1-line at $\Delta^2 = ln\ 4$ and then goes to the minimum value

$$\left[1 + \frac{2}{\pi}K_n^2\right]^{-1} \quad \text{as } \Delta^2 \to \infty. \tag{6.15}$$

Thus, the loss of efficiency of $\hat{\beta}_1^{\text{S}}$ relative to $\tilde{\beta}_1$ is

$$1 - \left[1 + \frac{2}{\pi}K_n^2\right]^{-1} \tag{6.16}$$

while the gain in efficiency is

$$\left[1 - \frac{2}{\pi}K_n^2\right]^{-1} \tag{6.17}$$

respectively which is achieved at $\Delta^2 = 0$. Thus, for $\Delta^2 < ln\ 4$, $\hat{\beta}_1^{\text{S}}$ performs better than $\tilde{\beta}_1$, otherwise $\tilde{\beta}_1$ performs better. The property of $\hat{\beta}_1^{\text{S}}$ is similar to the preliminary test estimator but does not depend on the level of significance. As $\Delta^2 \to \infty$ the relative

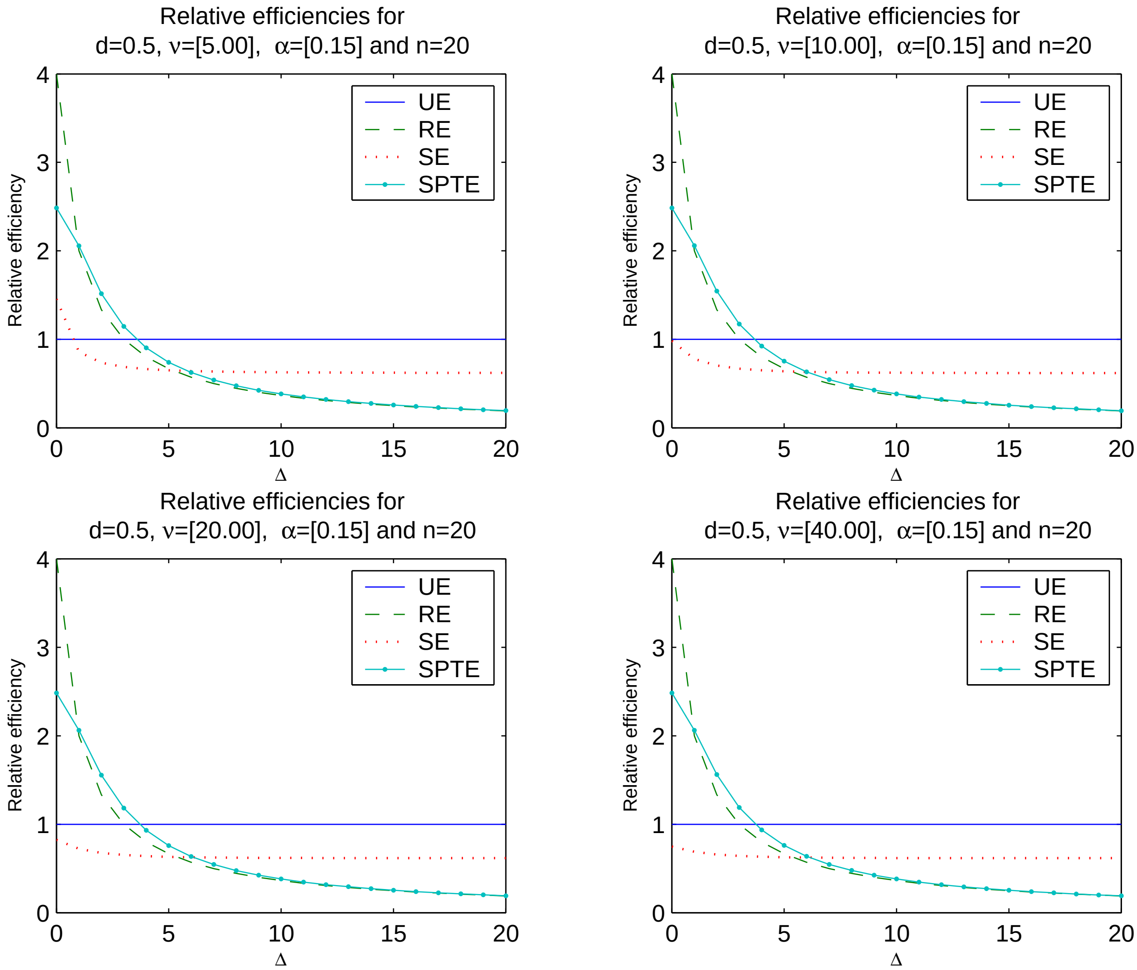


Figure 2: Graph of relative efficiency of RE, SPTE and SE relative to UE for $d = 0.5$ and different $\nu$

efficiency of SPTE with respect to UE approaches to 1 and that of the SE with respect to UE approaches to $\left[1 + \frac{2}{\pi} K_n^2\right]^{-1}$.

**Comparing efficiency of SE relative to SPTE**

The maximum relative efficiency of the SE relative to SPTE attained for $\Delta = 0$ and $d = 1$, regardless of the value of $\alpha$. At $\Delta = 0$, as the coefficient of distrust, $d$ decreases, the relative efficiency of SE also decreases, and it decreases below 1 for $d = 0$. Starting from some moderate value of $\Delta$, relative efficiency of SE becomes less than 1 and converges to a stable value, below one, as $\Delta \to \infty$. Except for $\Delta = 0$ and near 0 the relative efficiency of SE is always higher for smaller values of $d$ than larger values of $d$, before converging to a stable value. The difference between the relative efficiencies of SE for different values of $d$ is higher for lower value of $\alpha$ then it's higher values. As $\alpha$ increases this difference decreases. Moreover, as $\alpha$ increases, the relative efficiency of SE also increases for $\Delta = 0$

or near 0.

## 7 Concluding Remarks

The UE is based on the sample data alone and it is the only unbiased estimator among the four estimators considered in this paper. The introduction of the non-sample information in the estimation process causes the estimators to be biased. However, the biased estimators perform better than the unbiased estimator when they are judged based on the mse criterion. The performance of the biased estimators depend on the value of the departure parameter $\Delta$. In case of the SPTE, the performance also depends on the value of the level of significance. Under the null hypothesis, the departure parameter is zero, and the SE dominates all other estimators if $\alpha$ is not too high. As $\alpha$ increases, the performance of the SPTE improves when $\Delta$ is not too close to zero. At a lower level of significance, the SE performs better than the SPTE more often and over a wider range of values of $\Delta$. When the value of $\Delta$ is not far from 0, the SE always over performs the SPTE and RE. Therefore, in practice if the researcher could gather a value of $\beta_1$ from the prior knowledge or experience that is not too far from its true value, the SE would be the best choice as an 'improved' estimator of the slope.

**ACKNOWLEDGEMENTS**

The authors thankfully acknowledge some valuable suggestions of two unknown reviewers that have contributed to the improvement of the quality and presentation of the paper.

### References

Anderson, T.W. (1993). Nonnormal multivariate distributions: Inference based on elliptically contoured distributions. In Multivariate Analysis: Future Directions. ed by C.R. Rao. North-Holland, Amsterdam, 1–24.

Bancroft, T.A. (1944). On biases in estimation due to the use of the preliminary tests of significance. Ann. Math. Statist., **15**, 190–204.

Bolfarine, H. and Zacks, S. (1992). Prediction Theory for Finite Populations. Springer-Verlag, New York.

Box, G.E.P. and Tiao, G.C. (1992). Bayesian Inference in Statistical Analysis. Wiley, New York.

Celler, D., Fourdriunier, D and Strawderman, W.E. (1995). Shrinkage positive rule estimators for spherically symmetric distributions. *Journal of Multivariate Analysis*, **53**, 194-209.

Chmielewski, M.A. (1981). Elliptically symmetric distributions. *International Statistical Review*, Vol. 49, 67–74.

Chiou, P. and Saleh, A.K.Md.E. (2002). Preliminary test confidence sets for the mean of a multivariate normal distribution. Journal of Propagation in Probability and Statistics, **2**, 177-189.

Fang, K.T. and Anderson, T.W. (1990). Statistical Inference in Elliptically Contoured and Related Distributions. Allerton Press Inc., New York.

Fang, K.T. and Zhang, Y. (1990). Generalized Multivariate Analysis, Science Press and Springer-Verlag, New York

Fisher, R.A. (1956). Statistical Methods in Scientific Inference. Oliver and Boyd, London.

Fisher, R.A. (1960). The Design of Experiments. 7th ed., Hafner, New York.

Fraser, D.A.S. and Fick, G.H. (1975). Necessary analysis and its implementation. In Proc. Symposium on Statistics and Related Topics. ed. by A.K. Md. E. Saleh, Carleton University, Canada, 5.01–5.30.

Fraser, D.A.S. (1979). Inference and Linear Models, McGraw–Hill, New York.

Giles, J.A. (1991). Pre-testing for linear restrictions model with spherically symmetric disturbances. Journal of Econometrics, Vol. 50, 377–398.

Han, C.P. an Bancroft, T.A. (1968). On pooling means when variance is unknown. Jou. Amer. Statist. Assoc., **63**, 1333-1342.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, California, 361–379.

Judge, G.C. and Bock, M.E. (1978). Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics. North Holland, Amsterdam.

Khan, S. (1997). Likelihood based inference on the mean vectors of two multivariate Student-t populations. Far East Journal of Theoretical Statistics. 1, 1-17.

Khan, S. and Haq, M.S. (1990). Prediction distribution for the linear model with multivariate Student-t error. *Communications in Statistics: Theory & Methods*, Vol. 19, 4705–4712.

Khan, S. and Saleh, A.K.Md.E. (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. Biometrical Journal. Vol. 39, 131–147.

Khan, S. and Saleh, A.K.Md.E. (1998). Comparison of estimators of the mean based on $p$-samples from multivariate Student-t population. Communications in Statistics: Theory & Methods. Vol. 27(1), 193-210.

Khan, S. and Saleh, A.K.Md.E. (2001). On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. Statistical Papers, **42**(4), 451-473.

Lange, K.L. Little, R.J.A., and Taylor, J.M.G. (1989). Robust statistical modeling using the $t$-distribution. Journal of American Statistical Association, Vol. 84, 881–896.

Lobato, I. N. (2001). Testing That a Dependent Process Is Uncorrelated, Journal of the American Statistical Association, Volume 96(455), 1066-1076(11).

Maatta, J.M. and Casella, G. (1990). Developments indecision-theoretic variance estimation, *Statistical Sciences*, Vol. 5, 90–101.

Prucha I.R. and Kelegian, H.H. (1984). The structure of simultaneous equation estimators: A generalization towards non-normal disturbances, Econometrica, Vol. 52, 721–736.

Saleh, A K Md E. (2006). Theory of Preliminary Test and Stein-Type Estimation with Applications. Wiley, New York.

Saleh, A.K.Md.E. and Sen, P.K. (1978). Non-parametric estimation of location parameter after a preliminary test on regression. *Annals of Statistics,* **6**, 154-168.

Sclove, S.L., Morris, C. and Radhakrishnan, R. (1972). Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. *Ann. Math. Statist.*, Vol. 43, 1481–1490.

Sen, P.K and Saleh, A.K.Md.E. (1985). On some shrinkage estimators of multivariate location. *Annals of Statistics.* **13**, 272-281.

Singh, R.S. (1988). Estimation of error variance in linear regression models with errors having multivariate Student-t distribution with unknown degrees of freedom. Economics Letters, Vol. 27, 47–53.

Spanos, A. (1994). On modeling heteroskedasticity: The Student's t and elliptical linear regression models. Econometric Theory, Vol. 10, 286–315.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Math. Statist. and Probability, University of California Press, Berkeley, Vol. 1, 197-206.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. **9**, 1135–1151.

Ullah, A. and Walsh, V.Z. (1984). On the robustness of LM, LR and W tests in regression models. Econometrica, Vol. 52, 1055–1066.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. Journal of American Statistical Association, Vol. 66, 601–616.